KINECT Activity Detection in a Collaborative Environment

Faisal Ahmed and Marina Gavrilova

Department of Computer Science, University of Calgary, Calgary, AB, Canada
{faahmed, mgavrilo}@ucalgary.ca

Abstract

Traditionally, spatial data and knowledge was studied in a global context. Cities, countries, and geographical features could be observed, mapped, plotted, and modeled using an array of GIS tools. More and more nowadays, individual activities as well as spatio-temporal behavioral patterns of individuals in both online and real world are being studied, modeled, and observed. One of such environments is human spatial location tracking and activity analysis in collaborative environments. This work presents a new approach for using KINECT for activity recognition in a meeting room.

Background

Meetings are one example of such an environment, which facilitates exchange and distribution of information as well as creation and sharing of knowledge (Yu et al., 2007). However, a large amount of the total time spent in meetings is reported to be spent on non-essential tasks (Mosvick & Nelson, 1987). This is due to the inefficient login process, distortion of information, resource access and permissions, sub-optimal decision making, and mismanagement (Mosvick & Nelson, 1987). There has been a growing demand from industries for new generations of interactive technologies to support high productivity and to optimize time spent during meetings. In addition, secure real-time user authentication and access right management solutions are being actively researched and deployed in both universities and corporate settings (Deutschmann et al., 2013; Penteado & Marana, 2006; Zhou & Bhanu, 2007). In response to these demands, researchers are actively seeking new communication and collaboration technologies, which will reduce meeting attendees' enrollment time and provide the meeting organizer with the tools for efficient real-time attendance tracking and access resource management.

Traditional approaches to the problem involve the use of IDs, smart cards, user password/logon, coupled with user identity management solutions, all of which require a significant amount of time and additional resources (Kumar, 2010; Jeon et al., 2011). Rapid development of new technologies such as KINECT opened the door to a new class of fast and reliable identity management solutions, and changed the research landscape (Penteado & Marana, 2006; Down & Sands, 2004; Yanushkevich et al., 2007). A biometric system is a pattern recognition system that can recognize individuals based on their appearance or behavior. We base our research on the premises that, physiological and behavioral biometrics can be seamlessly integrated with technologies enabling meeting room setup, and, in addition to individual access management, can provide highly efficient group authentication capabilities. The objective is to enable meeting organizer to immediately start the meeting, keep track of attendance, and allow all participants to seamlessly access shared resources without compromising secure contents. In addition, the developed system should support archiving meeting statistics, determining input of each of the participants, and recognizing certain activities. This, in turn, will help to create more conductive collaborative environment, use time more effectively, and inform project managers on successful patterns.

In this presentation, we describe research in progress towards developing an activity recognition KINECT based system during meetings. A meeting room represents a dynamic collaborative environment, where different individual and group interactions take place. Accurate identification and summarization of these activities and interactions require reliable acquisition, synthesis, analysis, and integration of multiple sources of data. The objective is to present a discussion of the underlying technologies and the key design issues of a visual KINECT based meeting room system. We try to address the following open questions, which are imperative for the realization of an intelligent collaborative environment:

- How to perform individual visual identification and authentication in an effective manner?
- How to effectively analyze the meeting room workflow? How to identify the different activities among the participants?
- How to evaluate individual contribution level? Which features are required to be considered?
- How to visually analyze group dynamics and behaviors? Which features are important to be considered for this analysis?

Methodology

Our proposed multi-modal system comprises four major components: i) multi-modal sensing, ii) individual identification and tracking, iii) activity recognition, and iv) individual contribution analysis. Figure 1 shows the architecture, the components, and the workflow of our proposed system. We present our current research efforts toward building a multi-modal biometric-based authentication and activity detection system for collaborative environment. We propose to integrate complementary modalities, such as audio and video for accurate individual recognition and authentication. In addition, we present a Kinect-based 3D visual activity tracking system that tracks different activities based on Kinect skeleton data. In addition, audio-based activities can be tracked from the data collected by the Kinect microphone-array. These tracked activities are then used to define individual contribution and collaboration level. Our proposed system does not rely on any extensive human-computer interactions, rather it processes multi-modal data sources for automated authentication and activity recognition. The concepts presented here can also be extended to design online and virtual collaborative meeting room environments, which can be used to monitor and track online interactions and contributions of meeting participants. Note that an extended version of this work has been submitted as a book chapter to KredibleNet Book “Socio-computational frameworks, tools, and algorithms for supporting transparent authorship in social media knowledge markets”.

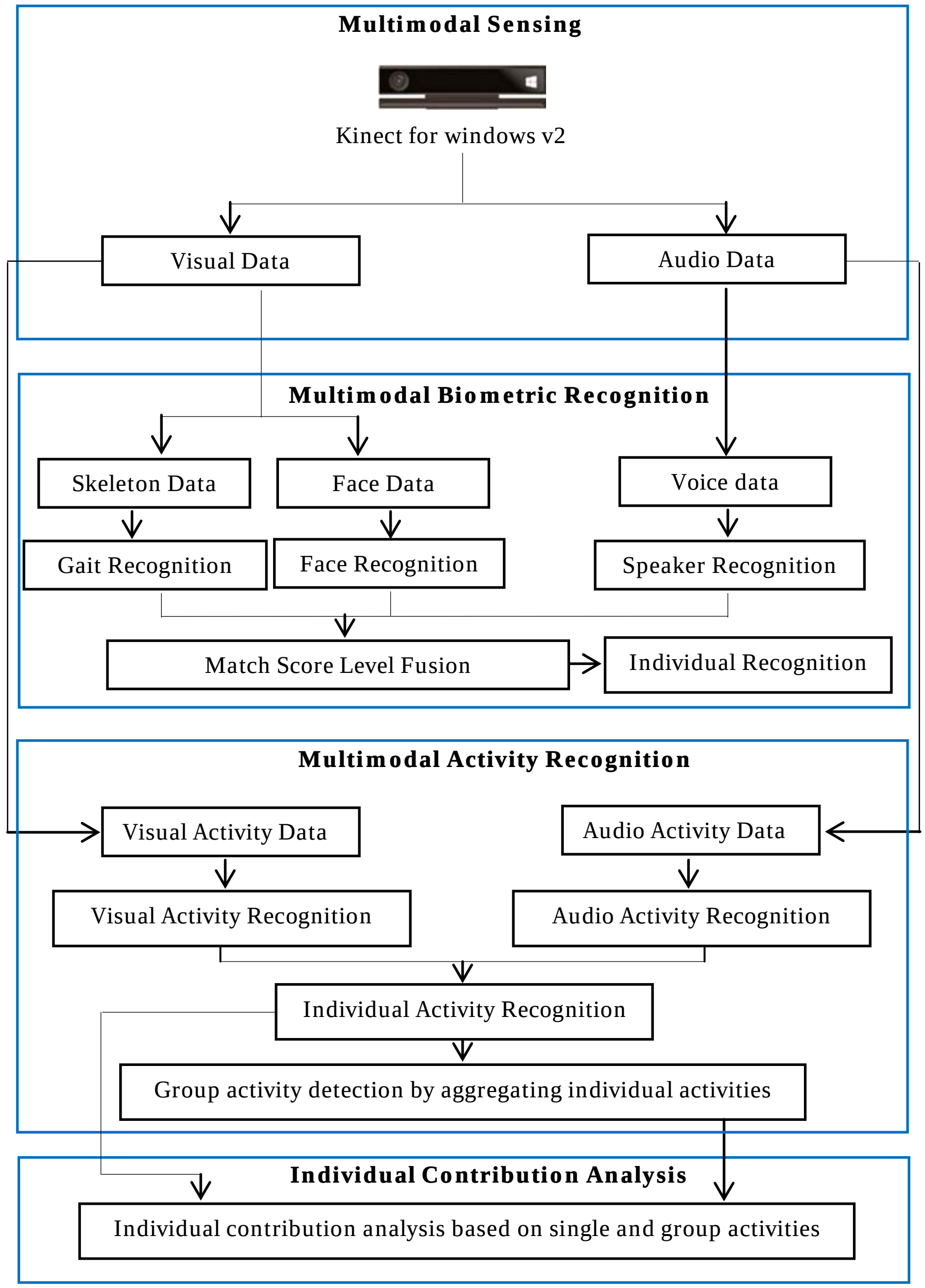

Fig. 1. Components of the proposed multi-modal system.

Activity recognition is one of the key components of any intelligent system designed for a collaborative environment. In our proposed system prototype, we analyze both video and audio data in order to recognize two types of activities: individual activities and group activities. In our proposed multi-modal system, we aim to detect 11 different activities. Table 1 lists the selected activities. Detection of these activities provides important context information regarding the meeting. In addition, it also facilitates analyzing participant contribution level. We propose to use the Kinect-based 3D skeleton model in order to detect activities like entering the room, exiting the room, sitting down, and getting up. The Kinect skeleton model provides skeleton point positions in a 3D space, which is more robust than the 2D head tracker-based methods. In addition we intent to use local texture-based feature representation for detecting a facial expression change.

Detecting audio-based activities such as presenter speaking and non-presenter speaking can be performed using voice-print recognition. We define presenter as any individual who is standing in front of the whiteboard or presentation board in the meeting room. On the other hand, non-presenter is any individual who is not standing in front of the presentation board.

Table 1. List of selected individual activities

Visual activities	Audio-based activities
1. Enter the room 2. Exit the room 3. Sit down 4. Get up 5. Raise hand 6. Gesture 7. Change of facial expression 8. Note-taking 9. Idle	1. Presenter speaking 2. Non-presenter speaking

Results

Based on the proposed activity detection approach, we have built a "sit" and "walk" activity tracker. The system tracks any user who enters the meeting room and shows the current status of the user ("walk/stand" or "sit") at the top of his head. In our experiments, we found that, the detection of walk and sit actions are quite accurate and consistent over the entire video sequence. We run some preliminary experiments where two subjects enter the room, sit on chairs for some time and then leave the room. Figure 2 illustrates the results for the two person activity recognition.

Acknowledgements

Authors would like to acknowledge NSERC DG, Alberta Ingenuity, NSERC ENGAGE, and SMART Technologies ULC for partial support of the project.

References

Deutschmann, I., Nordstrom, P., & Nilsson, L. (2013) Continuous Authentication Using Behavioral Biometrics. *IT Professional*, 15(4), 12-15.

Down, M., & Sands, R. (2004). Biometrics: An overview of the Technology, Challenges and Control Considerations. *Journal of Information Systems Control*, 4, 53-56.

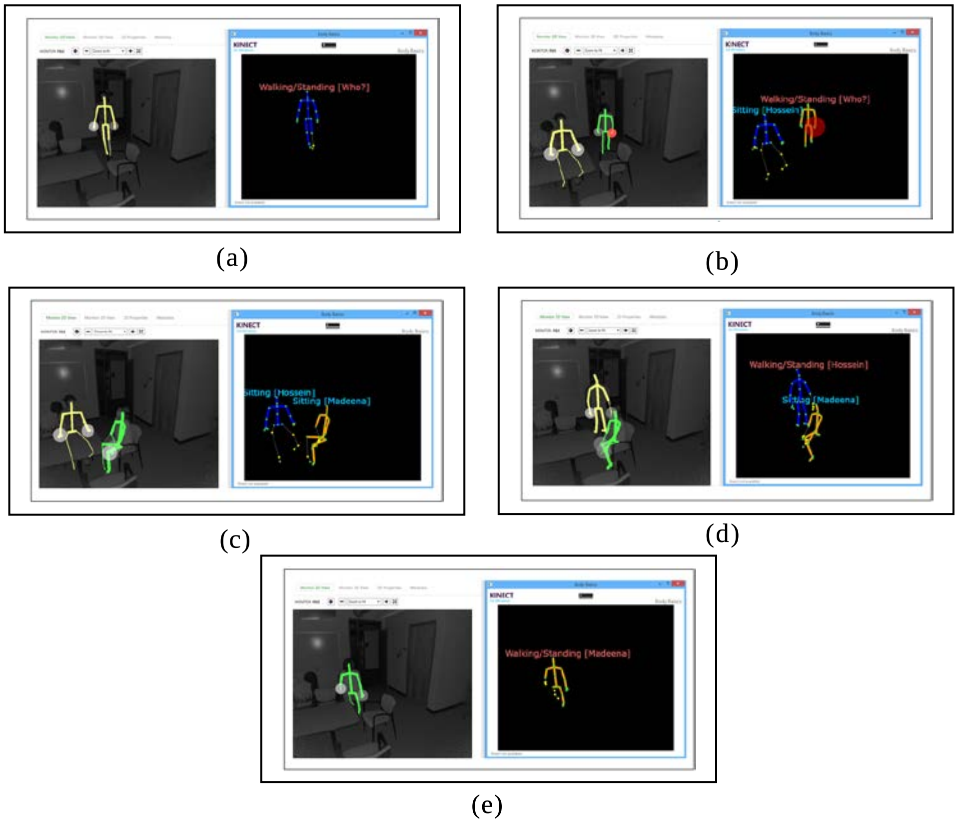

Fig. 2. Activity tracking (a) the first person enters the room (b) the first person sits on a chair, the second person enters the room, (c) both are sitting on the chair, (d) the first person stands up and leaves the room (e) the second person leaves the room.

Jeon, I.S., Kim, H.S., & Kim, M.S. (2011). Enhanced Biometrics-based Remote User Authentication Scheme Using Smart Cards. *Journal of Security Engineering*, 8(2), 237-254.

Kumar, M. (2010). An Enhanced Remote User Authentication Scheme with Smart Card. *International Journal of Network Security*, 10(3): 175-184.

Mosvick, R., & Nelson, R. (1987). *We've got to start meeting like this! a guide to successful business meeting management*. Glenview, IL: Scott, Foresman.

Penteado, B.E., & Marana, A.N. (2009) A Video-Based Biometric Authentication for e-Learning Web Applications. *Enterprise Information Systems*, Springer Berlin Heidelberg, 770-779.

Yanushkevich, S., Gavrilova, M., Wang, P., & Srihari, S. (2007). Image pattern recognition: Synthesis & analysis in biometrics. *Series in Machine Perception & Artificial Intelligence*, 67.

Yu, Z., Ozeki, M., Fujii, Y., & Nakamura, Y. (2007). Towards Smart Meeting: Enabling Technologies and a Real-World Application. *Proc. ACM Intl. Conf. on Multimodal Interfaces*, 86-93.

Zhou, X., & Bhanu, B. (2007). Integrating face and gait for human recognition at a distance in video. *IEEE Transactions on Systems, Man, and Cybernetics Part B*, 37(5), 1119-1137.